INFORMATION AND CONTROL 38, 135–147 (1978)

A Formulation of Fuzzy Decision Problems with Fuzzy Information using Probability Measures of Fuzzy Events

T. OKUDA,* H. TANAKA,† AND K. ASAI†

* *Department of Management Engineering, Osaka Institute of Technology, Omiya, Asahiku, Osaka 535, Japan and*
† *Department of Industrial Engineering, University of Osaka Prefecture, Sakai, Osaka 591, Japan*

For decision problems in the real world, states of nature, information, and actions should be viewed as fuzzy events. The application of the fuzzy sets theory and the statistical decision theory to the decision problems in fuzzy events leads to a specific formulation of fuzzy decision problems and the definitions of entropy, worth of information, and quantity of information. Some results which are analogous to those in the statistical decision theory are given in this paper.

1. INTRODUCTION

Although the theory of decision problems with probabilistic uncertainty has been studied in detail by DeGroot (1970) and by Kullback (1959), much of the decision-making in the real world takes place in a fuzzy environment. The present paper deals with a fuzzy decision problem in which we can regard states of nature, feasible actions, and available information as fuzzy events given by Zadeh (1968). An example of fuzzy states may be described by such expressions as "It will be warm," "It will be cool," etc. This fuzziness stems from such adjectives as "warm," "cool," etc. Since fuzzy states are defined on a probability space, the decision problems in fuzzy events are related to both fuzziness and randomness.

The application of the fuzzy sets theory and the statistical decision theory to the decision problems in fuzzy events leads to a specific formulation of fuzzy decision problems and to the definitions of entropy, worth of information, and quantity of information. From these definitions this paper explains some results concerning perfect, probabilistic, and fuzzy information, which are analogous to those in the statistical decision theory. In other words, we have tried to extend some of the statistical decision theory to decision problems involving fuzzy events in this paper. Although a reason of considering the decision problems involving fuzzy events essentially lies in advantages similar to those discussed by Zadeh (1965) and by Fung and Fu (1973), there are some specific advantages arising from our formulation of the decision problem.

0019-9958/78/0382–0135$02.00/0

2. Probability of Fuzzy Events

Let S and S' be sets of events $\{s_1, ..., s_n\}$ and $\{s'_1, ..., s'_m\}$ with probabilities $\xi(s_i)$ and $\xi'(s'_j)$, respectively. Fuzzy events A and B are fuzzy sets on S and S' whose membership functions μ_A and μ_B are Borel mesurable. Let $\zeta(s_i, s'_j)$ be a joint probability of s_i and s'_j. The following relations are given by Zadeh (1968).

$$P(A) = \sum_i \mu_A(s_i)\, \xi(s_i),$$

$$P(A, B) = \sum_i \sum_j \mu_A(s_i)\, \mu_B(s'_j)\, \zeta(s_i, s'_j),$$

$$P(A \mid s'_j) = P(A, s'_j)/\xi'(s'_j),$$

$$P(A \mid B) = P(A, B)/P(B),$$

where $P(A)$ is called the probability of fuzzy event A, $P(A, B)$ is called the joint probability of fuzzy events A and B, and so forth.

The following will be used later.

Lemma 1. (i) $\sum_j P(A, s'_j) = P(A), \sum_i (s_i, B) = P(B),$

(ii) $P(A) + P(\bar{A}) = 1,$

(iii) $P(A, B) + P(A, \bar{B}) = P(A), P(A, B) + P(\bar{A}, B) = P(B),$

(iv) $P(\bar{A}, \bar{B}) = 1 - P(A) - P(B) + P(A, B),$

where $\bar{A}$ stands for the complement of A characterized by the membership function $1 - \mu_A(\cdot)$.

3. Decision Problems in Fuzzy Events

A statistical decision problem is generally represented in the form of the ordered 4-tuple $\langle S, D, \xi, u' \rangle$ where $S = \{s_1, ..., s_n\}$ is a set of states of nature, $D = \{d_1, ..., d_p\}$ is a set of actions, $\xi(\cdot)$ is a probability density function on S, and $u'(\cdot, \cdot)$ is a utility function on $S \times D$.

Similarly, let us define a decision problem in fuzzy events as the ordered 5-tuple $\langle \mathscr{F}, \mathscr{A}, \xi, u, g \rangle$ where $\mathscr{F} = \{F_1, ..., F_r\}$ is a set of fuzzy states which are fuzzy events on S, $\mathscr{A} = \{A_1, ..., A_q\}$ is a set of fuzzy actions which are fuzzy events on D, $u(\cdot, \cdot)$ is a utility function on $\mathscr{A} \times \mathscr{F}$, and $g(\cdot)$ is a function mapping from a set D to $[0, 1]$ which represents a preference ordering over a set of actions such that the larger $g(d_i)$ is, the more preferable d_i is. It should noted that regardless of the value of utility function, $g(d_i)$ is determined by an individual subjective judgement.

In view of the definition of probability of fuzzy events,

$$G(A_i) = \sum_k \mu_{A_i}(d_k)\, g(d_k), \qquad \text{for} \quad A_i \in \mathscr{A}$$

may represent a preference ordering over a set of fuzzy actions $\mathscr{A}$. The function $G(\cdot)$ is normalized as follows:

$$P_g(A_i) = G(A_i)\Big/\sum_k G(A_k), \qquad \text{for} \quad A_i \in \mathscr{A}.$$

The function $P_g(\cdot)$ can be regarded as a subjective measure for the preference order of actions, which is assumed to be independently of the value of utility function.

First of all, let us assume that $\mathscr{F}$ and $u(\cdot, \cdot)$ are specified in the following manner:

(i) There exists a partition on $\mathscr{F}$ such that $\mathscr{F}_1 \cup \cdots \cup \mathscr{F}_l = \mathscr{F}$, $\mathscr{F}_i \cap \mathscr{F}_j = \varnothing$ for $i \neq j$ and each $\mathscr{F}_i$ is orthogonal. Here, $\mathscr{F}_i$ is orthogonal if and only if $\sum_{F_j \in \mathscr{F}_i} \mu_{F_j}(s_k) = 1$ for all $s_k \in S$.

(ii) Each utility function $u_i(\cdot, \cdot)$ is defined on $\mathscr{A} \times \mathscr{F}_i$ and a total utility function $u(\cdot, \cdot)$ is the sum of utility functions $\{u_i(\cdot, \cdot)\}$.

As such an example that the assumptions are satisfied, consider the problem in which we have to decide whether we order an article or not. In the present case, for instance, $\mathscr{F} = \{\mathscr{F}_1, \mathscr{F}_2, \mathscr{F}_3\}$ may be expressed by such classes as $\mathscr{F}_1 =$ {good quality, sufficient, poor}, $\mathscr{F}_2 =$ {good design, sufficient, poor}, and $\mathscr{F}_3 =$ {short time for delivery, ordinary, long}. The words "good quality," "good design," and "short time for delivery" may belong to essentially different classes. Taking account of classes of adjectives {good, sufficient, poor} and {short, ordinary, long}, it will be possible to define the fuzzy sets in each $\mathscr{F}_i$ such that the orthogonal condition is satisfied. Thus it should be noted that a family of sets $\{\mathscr{F}_i\}$ is a partition of $\mathscr{F}$ and each $\mathscr{F}_i$ forms a partition on S in a sence such that $\mathscr{F}_i$ is orthogonal. Furthermore it is reasonable from the above to define the total utility as the sum of utility functions $\{u_i(\cdot, \cdot)\}$.

Each $u_i(\cdot, \cdot)$ can be obviously replaced with a utility function $u(\cdot, \cdot)$ on $\mathscr{A} \times \mathscr{F}$ such that $u(A_k, F_j) = u_i(A_k, F_j)$ for $F_j \in \mathscr{F}_i$. In what follows, we use $u(\cdot, \cdot)$ instead of $u_i(\cdot, \cdot)$ for simple notation.

The above assumptions suggest the following definition.

DEFINITION 1. A total utility of fuzzy action A_i is

$$U(A_i) = \sum P_g(A_i)\, u(A_i, F_j)\, P(F_j).$$

An optimal decision can be defined as a fuzzy action A_0 which maximizes $U(A_i)$. That is,

$$U(A_0) \equiv \max_i U(A_i).$$

Note that the optimal decision A_0 is a fuzzy set.

Let $X = \{x_1, ..., x_m\}$ be a message space. Throughout this paper it is assumed that a conditional probability $f(x_j \mid s_i)$ is known. From Bayes' Formula the posterior probability can be written by

$$\xi(s_i \mid x_j) = f(x_j \mid s_i)\, \xi(s_i)/f(x_j),$$

where

$$f(x_j) = \sum_i f(x_j \mid s_i)\, \xi(s_i).$$

With $\xi(s_i \mid x_j)$, the total utility of A_i given a message x_j is

$$U(A_i \mid x_j) = \sum_k P_g(A_i)\, u(A_i, F_k)\, P(F_k \mid x_j),$$

where $P(F_k \mid x_j)$ is the conditional probability of F_k given x_j. Thus the optimal decision A_{x_j} can be defined by

$$U(A_{x_j} \mid x_j) \equiv \max_i U(A_i \mid x_j).$$

Let us define probabilistic information e as observing the value of a random variable $\tilde{x}$. The total utility for having the information e is

$$U(A_{\tilde{x}} \mid \tilde{x}) = \sum_j U(A_{x_j} \mid x_j)\, f(x_j).$$

It is natural from the statistical decision theory that the worth of information e is written in the form

$$V(e) = U(A_{\tilde{x}} \mid \tilde{x}) - U(A_0).$$

Note that the worth of information $V(e)$ has been estimated before the information e is obtained.

Now consider the information which let us exactly know a true state s_i. This information will be called probabilistic perfect information and denoted by e_∞. If the information e_∞ is available, its total utility for having the information e_∞ is

$$U(A_{\tilde{s}} \mid \tilde{s}) = \sum_k U(A_{s_k} \mid s_k)\, \xi(s_k),$$

where

$$U(A_{s_k} \mid s_k) \equiv \max_i U(A_i \mid s_k)$$

and

$$U(A_i \mid s_k) = \sum_j P_g(A_i)\, u(A_i, F_j)\, \mu_{F_j}(s_k).$$

Thus the worth of information e_∞ can be similarly written by

$$V(e_\infty) = U(A_{\tilde{s}} \mid \tilde{s}) - U(A_0).$$

It is easy from these definitions to prove the following proposition.

PROPOSITION 1. $V(e_\infty) \geqslant V(e) \geqslant 0$.

Next we consider a fuzzy message M which is a fuzzy event on X. Let us define fuzzy information E as observing a fuzzy message from $\{M, \overline{M}\}$. From the same point of view the total utility of A_i given M is

$$U(A_i \mid M) = \sum_j P_g(A_i)\, u(A_i, F_j)\, P(F_j \mid M),$$

where $P(F_j \mid M)$ is the conditional probability of F_j given M, and the optimal decision A_M can be given by

$$U(A_M \mid M) \equiv \max_i U(A_i \mid M).$$

Then the total utility for having the fuzzy information E can be formulated as

$$U(A_{\tilde{M}} \mid \tilde{M}) = U(A_M \mid M)\, P(M) + U(A_{\overline{M}} \mid \overline{M})\, P(\overline{M})$$

which corresponds to that of the probabilistic information e.

DEFINITION 2. The worth of fuzzy information E is

$$V(E) = U(A_{\tilde{M}} \mid \tilde{M}) - U(A_0).$$

PROPOSITION 2. *The worth of fuzzy information E is nonnegative*; $V(E) \geqslant 0$. *If μ_M is constant or if $\tilde{x}$ and $\tilde{s}$ are mutually independent, then the equality holds.*

Proof. From Lemma 1, we have

$$\begin{aligned} U(A_{\tilde{M}} \mid \tilde{M}) &= U(A_M \mid M)\, P(M) + U(A_{\overline{M}} \mid \overline{M})\, P(\overline{M}) \\ &\geqslant \max_i \sum_j P_g(A_i)\, u(A_i, F_j)\{P(F_j, M) + P(F_j, \overline{M})\} \\ &= U(A_0). \end{aligned}$$

If μ_M is constant or if $\tilde{x}$ and $\tilde{s}$ are mutually independent, then $P(F_j \mid M) = P(F_j \mid \overline{M}) = P(F_j)$. Hence the equality holds. ∎

In fuzzy information, we consider the information which let us know what states $\{F_j\}$ occur with probability 1. This information will be called fuzzy perfect information E_∞ which corresponds to the probabilistic perfect information e_∞. To avoid the complexity of proofs we, however, limit ourselves to the simple case that given fuzzy perfect information E_∞ implies that $P(F_j) = 1$ and $P(F_k) = 0$ for all $k \neq j$. Even if E_∞ implies that $P(F_1) = \cdots = P(F_j) = 1$ and $P(F_k) = 0$ for $k > j$, it will be intuitively understood from the limited case that the following propositions still are satisfied.

In view of the above, the utility of A_i can be represented as $U(A_i \mid F_j) = P_g(A_i)\, u(A_i, F_j)$. The optimal decision A_{F_j} can be defined by

$$U(A_{F_j} \mid F_j) \equiv \max_i U(A_i \mid F_j).$$

Thus the total utility for having the fuzzy perfect information E_∞ is

$$U(A_{E_\infty} \mid E_\infty) = \sum_j U(A_{F_j} \mid F_j)\, P(F_j).$$

DEFINITION 3. The worth of E_∞ is

$$V(E_\infty) = U(A_{E_\infty} \mid E_\infty) - U(A_0).$$

PROPOSITION 3. $V(E_\infty) \geqslant V(E)$.

This proposition can be easily derived from Definitions 2 and 3.

PROPOSITION 4. (i) $V(E_\infty) \geqslant V(e_\infty)$,
(ii) $V(e) \geqslant V(E)$.

Proof. The proof of (i) is evident from definitions of $V(E_\infty)$ and $V(e_\infty)$. Hence we will show you only the proof of (ii).

$$\begin{aligned}
U(A_{\tilde{M}} \mid \tilde{M}) &= \max_i \sum_j P_g(A_i)\, u(A_i, F_j) \Big(\sum_k P(F_j \mid x_k)\, \mu_M(x_k)\, f(x_k)\Big) \\
&\quad + \max_i \sum_j P_g(A_i)\, u(A_i, F_j) \Big(\sum_k P(F_j \mid x_k)(1 - \mu_M(x_k))\, f(x_k)\Big) \\
&\leqslant \sum_k \Big\{\max_i \sum_j P_g(A_i)\, u(A_i, F_j)\, P(F_j \mid x_k)\Big\}\, \mu_M(x_k)\, f(x_k) \\
&\quad + \sum_k \Big\{\max_i \sum_j P_g(A_i)\, u(A_i, F_j)\, P(F_j \mid x_k)\Big\}\, (1 - \mu_M(x_k))\, f(x_k) \\
&= \sum_k U(A_{x_k} \mid x_k)\, f(x_k) \\
&= U(A_{\tilde{x}} \mid \tilde{x}). \quad \blacksquare
\end{aligned}$$

Remark. From Propositions 1 to 4, the following relation eventually holds:

$$V(E_\infty) \geqslant V(e_\infty) \geqslant V(e) \geqslant V(E) \geqslant 0.$$

It might be said that the relation $V(e) \geqslant V(E)$ is caused by the fact that the information E has fuzziness in addition to randomness which the information e has. On the other hand, the relation $V(E_\infty) \geqslant V(e_\infty)$ is caused by the fact that our interest is not in S but in $\mathscr{F}$ which is a set of our concerned events on S. This result will be intuitionally agreed on by all.

4. Quantity of Information in Fuzzy Events

Let $-\log P(A)$ be a measure of uncertainty of fuzzy event A and let the expected value of $-\log P(A)$ be a entropy of fuzzy event A.

Definition 4. The entropy of fuzzy event A is

$$H(\tilde{A}) = -P(A) \log P(A) - P(\bar{A}) \log P(\bar{A}).$$

The conditional probability of A given a message $\tilde{x} = x_j$ becomes $P(A \mid x_j)$. Consequently, the conditional entropy of fuzzy event A given the information e can be represented as

$$H(\tilde{A} \mid \tilde{x}) = \sum_j H(\tilde{A} \mid x_j) f(x_j),$$

where

$$H(\tilde{A} \mid x_j) = -P(A \mid x_j) \log P(A \mid x_j) - P(\bar{A} \mid x_j) \log P(\bar{A} \mid x_j).$$

Similarly, the conditional entropy of fuzzy event A given the fuzzy information E can be represented as

$$H(\tilde{A} \mid \tilde{M}) = H(\tilde{A} \mid M) P(M) + H(\tilde{A} \mid \bar{M}) P(\bar{M}).$$

Proposition 5. (i) $H(\tilde{A}) \geqslant H(\tilde{A} \mid \tilde{x})$,

(ii) $H(\tilde{A} \mid \tilde{x}, \tilde{y}) \leqslant H(\tilde{A} \mid \tilde{x}) + H(\tilde{A} \mid \tilde{y})$.

Proof. It is easy to prove (i) from Lemma 1 by using the fact that $\log x \leqslant x - 1$ for $x > 0$. For (ii), suppose that $\zeta(\cdot, \cdot)$ is a joint probability density function $\tilde{x}$ and $\tilde{y}$. Then we have

$$\begin{aligned} H(\tilde{A} \mid \tilde{x}) = & -\sum_i \sum_j P(A \mid x_i, y_j)\, \zeta(x_i, y_j) \log P(A \mid x_i) \\ & -\sum_i \sum_j P(\bar{A} \mid x_i, y_j)\, \zeta(x_i, y_j) \log P(\bar{A} \mid x_i) \end{aligned}$$

and

$$H(\tilde{A} \mid \tilde{y}) = -\sum_i \sum_j P(A \mid x_i, y_j)\, \zeta(x_i, y_j) \log P(A \mid y_j)$$
$$- \sum_i \sum_j P(\bar{A} \mid x_i, y_j)\, \zeta(x_i, y_j) \log P(\bar{A} \mid y_j),$$

since

$$\sum_j P(A \mid x_i, y_j) f(y_j \mid x_i) = P(A \mid x_i)$$

and

$$\sum_i P(A \mid x_i, y_j) f(x_i \mid y_j) = P(A \mid y_j).$$

Using $\log x \leqslant x - 1$ for $x > 0$ and $P(A \mid x_i)\, P(A \mid y_j) + P(\bar{A} \mid x_i)\, P(\bar{A} \mid y_j) \leqslant 1$, it follows that

$$-P(A \mid x_i, y_j) \log P(A \mid x_i, y_j) - P(\bar{A} \mid x_i, y_j) \log P(\bar{A} \mid x_i, y_j)$$
$$\leqslant -P(A \mid x_i, y_j) \log P(A \mid x_i)\, P(A \mid y_j) - P(\bar{A} \mid x_i, y_j) \log P(\bar{A} \mid x_i)\, P(\bar{A} \mid y_j)$$

for all x_i and y_j. Hence, we obtain

$$\begin{aligned} H(\tilde{A} \mid \tilde{x}, \tilde{y}) &= -\sum_i \sum_j \zeta(x_i, y_j)\{P(A \mid x_i, y_j) \log P(A \mid x_i, y_j) \\ &\qquad + P(\bar{A} \mid x_i, y_j) \log P(\bar{A} \mid x_i, y_j)\} \\ &\leqslant -\sum_i \sum_j \zeta(x_i, y_j)\{P(A \mid x_i, y_j) \log P(A \mid x_i)\, P(A \mid y_j) \\ &\qquad + P(\bar{A} \mid x_i, y_j) \log P(\bar{A} \mid x_i)\, P(\bar{A} \mid y_j)\} \\ &= H(\tilde{A} \mid \tilde{x}) + H(\tilde{A} \mid \tilde{y}). \quad \blacksquare \end{aligned}$$

In correspondence to Proposition 5, we have the following proposition.

PROPOSITION 6. (i) $H(\tilde{A}) \geqslant H(\tilde{A} \mid \tilde{B})$,

(ii) $H(\tilde{A}, \tilde{B}) = H(\tilde{A} \mid \tilde{B}) + H(\tilde{B})$,

(iii) $H(\tilde{A}, \tilde{B}) \leqslant H(\tilde{A}) + H(\tilde{B})$,

(iv) $H(\tilde{A} \mid \tilde{B}, \tilde{C}) \leqslant H(\tilde{A} \mid \tilde{B}) + H(\tilde{A} \mid \tilde{C})$.

Proof.

$$
\begin{aligned}
\text{(i)}\quad & H(\tilde{A}) - H(\tilde{A} \mid \tilde{B}) \\
&= -P(A, B) \log \frac{P(A)}{P(A \mid B)} - P(A, \bar{B}) \log \frac{P(A)}{P(A \mid \bar{B})} \\
&\quad - P(\bar{A}, B) \log \frac{P(\bar{A})}{P(\bar{A} \mid B)} - P(\bar{A}, \bar{B}) \log \frac{P(\bar{A})}{P(\bar{A} \mid \bar{B})} \\
&\geqslant -P(A, B) \left\{ \frac{P(A)}{P(A \mid B)} - 1 \right\} - P(A, \bar{B}) \left\{ \frac{P(A)}{P(A \mid \bar{B})} - 1 \right\} \\
&\quad - P(\bar{A}, B) \left\{ \frac{P(\bar{A})}{P(\bar{A} \mid B)} - 1 \right\} - P(\bar{A}, \bar{B}) \left\{ \frac{P(\bar{A})}{P(\bar{A} \mid \bar{B})} - 1 \right\} \\
&= 0.
\end{aligned}
$$

(ii) We have the following relation

$$
\begin{aligned}
H(\tilde{A}, \tilde{B}) = &-P(A, B) \log P(A, B) - P(\bar{A}, B) \log P(\bar{A}, B) \\
&- P(A, \bar{B}) \log P(A, \bar{B}) - P(\bar{A}, \bar{B}) \log P(\bar{A}, \bar{B}),
\end{aligned}
$$

so that (ii) can be easily proved by using Lemma 1 from the definitions of $H(\tilde{A} \mid \tilde{B})$ and $H(\tilde{B})$.

(iii) This assertion follows from (i) and (ii).

(iv) In view of the proof of Proposition 5, it is clear that this assertion holds. ∎

On the basis of Definition 3 we have the following definitions.

DEFINITION 5. The entropy of a state space $\mathscr{F}$ is

$$H(\mathscr{F}) = -\sum_i \{P(F_i) \log P(F_i) + P(\bar{F}_i) \log P(\bar{F}_i)\}.$$

From Definition 5 the conditional entropy of $\mathscr{F}$ given the information e can be written by

$$H(\mathscr{F} \mid \tilde{x}) = \sum_j H(\mathscr{F} \mid x_j) f(x_j).$$

The conditional entropy of $\mathscr{F}$ given the fuzzy information E can be similarly written by

$$H(\mathscr{F} \mid \tilde{M}) = H(\mathscr{F} \mid M) P(M) + H(\mathscr{F} \mid \bar{M}) P(\bar{M}).$$

DEFINITION 6. The quantity of information e and the quantity of fuzzy information E are

$$I(e) = H(\mathscr{F}) - H(\mathscr{F} \mid \tilde{x})$$

and

$$I(E) = H(\mathscr{F}) - H(\mathscr{F} \mid \tilde{M})$$

respectively.

The following proposition can be immediately proved from Propositions 5 and 6.

PROPOSITION 7. (i) $H(\mathscr{F} \mid \tilde{x}, \tilde{y}) \leqslant H(\mathscr{F} \mid \tilde{x}) + H(\mathscr{F} \mid \tilde{y})$

(ii) $I(e) \geqslant 0$,

(iii) $H(\mathscr{F} \mid \tilde{M}_1, \tilde{M}_2) \leqslant H(\mathscr{F} \mid \tilde{M}_1) + H(\mathscr{F} \mid \tilde{M}_2)$

(iv) $I(E) \geqslant 0$.

PROPOSITION 8. *Suppose that there exist two fuzzy information* E_1 *and* E_2. *Then for* $\forall \xi$,

$$V(E_1) \geqslant V(E_2) \textit{ for any utility function} \Leftrightarrow I(E_1) \geqslant I(E_2).$$

The proof of this proposition is analogous to that in the statistical information theory. More exactly speaking, the following relation in a non-fuzzy case is well known;

$$\begin{aligned} &V(e_1) \geqslant V(e_2) \text{ for any utility function} \\ &\quad \Leftrightarrow E_x\{\psi(f(x)) \mid \xi\} \geqslant E_y\{\psi(f(y)) \mid \xi\} \qquad \text{for any convex function } \psi, \end{aligned}$$

where $e_1 = \{x_1, ..., x_n\}$ and $e_2 = \{y_1, ..., y_m\}$ are two probabilistic information sources, and E_x and E_y mean taking an average by probabilistic variables x and y, respectively. Since we can regard two fuzzy information sources $E_1 = \{M_1, \bar{M}_1\}$ and $E_2 = \{M_2, \bar{M}_2\}$ as kinds of probabilistic variables and the quantity of fuzzy information source $I(\cdot)$ is a convex function, we can easily obtain Proposition 8 in applying the above relation to a fuzzy case [for example, see Miyazawa's book (1971)].

We conclude this section with a few remarks concerning experiments. The word "experiment" is used here in a sence such that we obtain a message x_j.

First, letting $X_l = (x_1, ..., x_l)$ be a sequence of experiments, we introduce four sets $S_1 = \{s_k \in S \mid \xi(s_k \mid X_l) > 0\}$, $S_2 = \{s_k \in S \mid 0 < \mu_A(s_k) < 1\}$, $S_3 = \{s_k \in S \mid \mu_A(s_k) = 1\}$, and $S_4 = \{s_k \in S \mid \mu_A(s_k) = 0\}$.

Since $S_1 \subseteq S_3$ implies $P(A \mid X_l) = 1$ and $S_1 \subseteq S_4$ implies $P(A \mid X_l) = 0$, it follows that if $S_1 \subseteq S_3$ or $S_1 \subseteq S_4$, then $H(\tilde{A} \mid X_l) = 0$. Thus there is no uncertainty of fuzzy event A after we carried out the experiments X_l.

In the case that $S_1 \subseteq S_3$, the decision maker can be convinced that the fuzzy event A occurs surely. Figure 1a illustrates the situation in which it is unnecessary to perform additional experiments. In the case that $S_1 \cap S_2 \neq \varnothing$, it follows that $H(\tilde{A} \mid X_l) > 0$. Figure 1b illustrates such a situation in which it is possible to decrease the entropy by performing additional experiments.

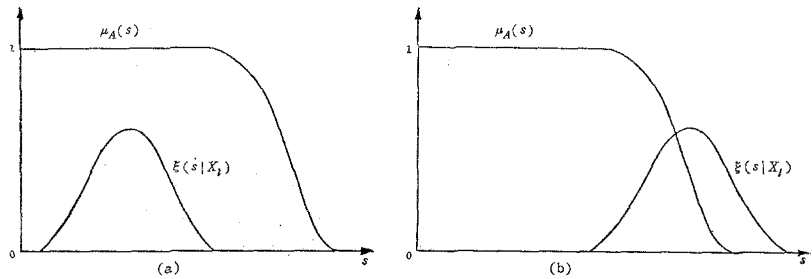

FIG. 1. Membership function $\mu_A(\cdot)$ and probability density function $\xi(\cdot \mid X_l)$ given X_l.

Next, we consider the case of Fig. 1b such that additional information is meaningful. It is known in the statistical decision theory that after infinite experiments X_∞ the entropy of events $\{s_i\}$ becomes zero.

On the other hand, it follows from the definition of entropy of fuzzy event A that

$$H(\tilde{A} \mid X_\infty) = -\mu_A(s_i) \log \mu_A(s_i) - (1 - \mu_A(s_i)) \log(1 - \mu_A(s_i)),$$

where s_i is a true state. Thus, the uncertainty expressed as $H(\tilde{A} \mid X_\infty) > 0$ still remains after infinite experiments X_∞ if $s_i \in S_2$. The entropy given X_∞ is similar to that without randomness in De Luca and Termini (1972). After infinite experiments X_∞ we see that the uncertainty of fuzzy event A does not arise from the randomness of $\{s_i\}$, but from the fuzziness of meaning of the word "A."

5. Concluding Remarks

We conclude this paper with summing up the meanings of this approach for formulating decision problems as follows:

(i) There are many actual decision problems such that it is necessary only to decide approximately what actions should be selected. An optimal fuzzy action A_0 obtained in this paper might show such a guiding principle in our actions.

(ii) It is reasonable that we consider the fuzzy state space $\mathscr{F}$ consisted of all F_i rather than the state space S, where each F_i is expressed by our interested words on S. For example, consider a decision problem in which we have to tell whether we approve of the plan to build up an atomic power plant. Let $S = \{s_i\}$ be a set of quantities of radioactive contamination from this plant. Even if we know that some state s_i occurs with probability one, our interest is only in whether the state s_i is safe for us or not. Therefore it is reasonable to analize this decision problem on the fuzzy state space $\mathscr{F} = \{F_1, F_2, F_3\} =$ {sufficiently safe, approximately safe, otherwise} rather than on S, where F_i is given by the orthogonal condition of $\mathscr{F}$.

(iii) Since most of information in the real world is fuzzy, we have more to do with fuzzy information rather than non-fuzzy information. Furthermore, there are many cases in which we must deal with fuzzy information even when we get non-fuzzy information. For example, even if we obtain the information which tells us that the quantity of a polluted material is x ppm, we must discuss how we interpret this x ppm. That leads to convert decisive information to fuzzy information. Thus it is necessary to view non-fuzzy information as fuzzy information defined by the meanings of non-fuzzy information.

(iv) Since there generally exists a large number of elements $|D| \times |S|$, it costs a great deal to define $u(d_i, s_j)$ for all i, j, where $|\ |$ denotes a number of elements in the set. The decision problems on $\mathscr{A} \times \mathscr{F}$ can be easily analyzed because of a smaller number of elements $|\mathscr{A}| \times |\mathscr{F}|$ than $|D| \times |S|$. Here a utility function $u(A_i, F_j)$ is subjectively defined on $\mathscr{A} \times \mathscr{F}$. Hence, we can easily decide an optimal decision A_0 from our formulation and then reanalyze the decision problem on a smaller subset of actions $A^* = \{d \mid \mu_{A_0}(d) \geqslant \alpha^*\}$, where α^* denotes a satisfactory threshold.

(v) Since the "experiment" is to obtain a message, it makes no sense to repeat many experiments at a great cost in the situation where the meanings of events are very fuzzy. From the degree of uncertainty of the "fuzzy" nature we, hence, can judge whether it is meaningful to repeat more experiments or not.

It follows from the above point of view that our intuitive concept of decision may be well reflected in our formulation of decision problems in a fuzzy environment.

RECEIVED: January 24, 1975; REVISED: September 16, 1977

References

DeGroot, M. H. (1970), "Optimal Statistical Decisions," McGraw–Hill, New York.
De Luca, A., and Termini, S. (1972), A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory, *Inform. Contr.* **20**, 301.
Fung, L., and Fu, K. S. (1973), Decision-Making in a Fuzzy Environment, TR-EE 73-22, School of Electrical Engineering, Purdue University.
Kullback, S. (1959), "Information Theory and Statistics," Wiley, New York.
Miyazawa, K. (1971), "Introduction to Information and Decision Theories," Iwanami Shoten, Tokyo (in Japanese).
Zadeh, L. A. (1965), Fuzzy sets, *Inform. Contr.* **8**, 338.
Zadeh, L. A. (1968), Probability measures of fuzzy events, *J. Math. Anal. Appl.* **23**, 421.